Science & Society

EMBO reports

The big health data sale

As the trade of personal health and medical data expands, it becomes necessary to improve legal frameworks for protecting patient anonymity, handling consent and ensuring the quality of data

Philip Hunter

Personal health and medical data are a valuable commodity for a number of sectors from public health agencies to academic researchers to pharmaceutical companies. Moreover, "big data" companies are increasingly interested in tapping into this resource. One such firm is Google, whose subsidiary Deep Mind was granted access to medical records on 1.6 million patients who had been treated at some time by three major hospitals in London, UK, in order to develop a diagnostic app. The public discussion it raised was just another sign of the long-going tensions between drug companies, privacy advocates, regulators, legislators, insurers and patients about privacy, consent, rights of access and ownership of medical data that is generated in pharmacies, hospitals and doctors' surgeries. In addition, the rapid growth of eHealth will add a boon of even more health data from mobile phones, portable diagnostic devices and other sources.

These developments are driving efforts to create a legal framework for protecting confidentiality, controlling communication and governing access rights to data. Existing data protection and human rights laws are being modified to account for personal medical and health data in parallel to the campaign for greater transparency and access to clinical trial data. Healthcare agencies in particular will have to revise their procedures for handling medical or research data that is associated with patients.

Google's foray into medical data demonstrates the key role of health agencies, in this case the Royal Free NHS Trust, which operates the three London hospitals that granted Deep Mind access to patient data. Royal Free approached Deep Mind with a request to develop an app for detecting acute kidney injury, which, according to the Trust, affects more than one in six inpatients. The Trust declined to comment and referred to a prepared statement (https://www.royalfree.nhs.uk/news-media/news/google-deepmind-qa/) pointing out that the app called Streams improves the detection of acute kidney injury by immediately reviewing blood test results for signs of deterioration and sending an alert and the results to the most appropriate clinician. A key benefit for health agencies such as Royal Free is getting access to sophisticated data mining and analysis to improve diagnostic tests and provide early warning of impending conditions.

> *"Existing data protection and human rights laws are being modified to account for personal medical and health data..."*

The Trust's statement addressed criticism for not informing patients that their data had been made available to an outside party by pointing out that the sharing agreement with Deep Mind conformed with UK legislation. All information sent to and processed by Deep Mind was encrypted both in transit and while being stored within the Deep Mind Health computing cluster. The Trust also cited the UK's Caldicott Information Governance Review (https://www.gov.uk/government/publications/the-information-governance-review), according to which health professionals may rely on implied consent when sharing personal data in the interests of direct care. The Trust pointed out that the NHS has data sharing agreements with a number of third party organizations, many of which, it argued, were vital to the safe and effective treatment of patients. It would not be practical or safe to ask every patient to consent to every one of these arrangements, the Trust's statement added.

Ross Anderson, Professor of Security Engineering at the University of Cambridge, UK, broadly supported the Trust's argument. "In the specific context of the Royal Free dataset, Deep Mind promised to keep the data confidential and to keep it in the UK, rather than putting it on Google's servers in the USA, and to use the data only to develop tools for direct patient care. So what they were doing was within the legal envelope set out by data protection law", he said. Anderson, who has been critical of current laws for the protection of personal medical data, suggested that Royal Free could have contacted patients to inform them of their right to opt out, but he was dismissive of arguments in the UK press that the records might leak out.

Google would not comment directly but Mustafa Suleyman, Co-Founder of Deep Mind, commented that, "[w]e are working with clinicians at the Royal Free to understand how technology can best help clinicians recognise patient deterioration—in this case acute kidney injury (AKI). We do, and will always, hold ourselves to the highest possible standards of patient data protection. This data will only ever be used for the purposes of improving health care and will never be linked with Google accounts or products".

However, the involvement of Deep Mind in processing personal medical data highlights an emerging concern for healthcare agencies, namely the power and leverage such companies might gain

Freelance journalist in London, UK. E-mail: ph@philiphunter.com
DOI 10.15252/embr.201642917 | Published online 11 July 2016

through their expertise. Anderson is concerned that it could create a monopolistic position for the analysis they provide, whether in diagnostics or recommendations for treatment. "This is something the NHS will have to start thinking about", he said. "The monopolies given to drug companies via the patent system led to abuses that had to be curtailed by pricing systems and the same is likely to be true of valuable insights derived from big data by AI (Artificial Intelligence) and machine learning techniques".

In fact, there is significant public distrust over companies that are ultimately motivated by profit. The key to winning public trust is to therefore convince people that the data is being used in their interests, according to a 2016 report by the Wellcome Trust (https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf). Their survey revealed that a slight majority (53%) of people would agree to their data being used by commercial organizations if it was for research, while there is widespread concern over insurance and marketing companies getting access to personal health data. A significant minority of respondents (17%) objected to giving private companies access to health data under any circumstances. "We believe patient data gathered within the healthcare system has tremendous potential for biomedical and health research", commented Beth Thompson, Senior Policy Adviser at the Wellcome Trust. "However, patients must have clear information about what happens to their data, how it may be used, and be given an opportunity to opt out".

> *"The key to winning public trust is to therefore convince people that the data is being used in their interests…"*

The report also sheds light on another risk, namely that the involvement of big companies in health care changes the context of patient consent. Patients generally accept giving information to doctors in return for better health care. Customers also accept giving some personal data to commercial companies in return for benefits such as discounts. The Wellcome report pointed out that these are two clearly established contexts for data sharing, but with different motivations and mindsets. "When commercial companies are involved in the health service and in health research, the distinction between these contexts collapses", according to the report. "Unsure whether we are using a service or making a transaction, we find it harder to assess the risks and benefits of our data being made accessible. In these situations, participants in the research were often cautious about the idea of commercial access to their data but struggled to articulate why; they simply had an intuitive sense of discomfort. This suggests traditional norms and paradigms are being challenged, with uncertain consequences".

> *"… the involvement of big companies in health care changes the context of patient consent"*

This "contextual collapse" is already underway in the USA, driven by a fast growing trade in personal health data. Walgreens and CVS, the country's two largest pharmacy chains, already acquire and trade in personal medical information obtained by offering incentives such as discounts and free offers. Walgreens offers an app for smartphones that monitors blood glucose and blood pressure and feeds the data into the chain's network in exchange for discounts. Meanwhile, CVS is offering its customers a US$5 rebate for every 10 prescription refills in return for waiving their rights to privacy under the federal Health Insurance Portability and Accountability Act (HIPSAA). Passed in 1996, HIPSAA governs transfer of healthcare data tied directly to an individual's identity.

Although medical data must be anonymized, it is a valuable resource for drug companies; Pfizer is on record as stating that it spends US$12 million a year acquiring anonymized health data (http://www.scientificamerican.com/article/how-data-brokers-make-money-off-your-medical-records/). One source of medical data is IMS Health, the world's largest player in medical data trading with revenues upwards of US$2.9 billion in 2015. Its closest rival is Symphony Health Solutions, and the practice is extending around the world, although often subject to tighter privacy restrictions outside the USA.

While Walgreens and CVS declined requests for interviews, the data specialists were more open about addressing issues of privacy and anonymity. Heather Varela, Symphony's Director for Marketing Communications, stressed that it employed many measures to protect the anonymity of data held in its repositories and comply with HIPAA requirements. In addition, Symphony also recognizes that its own employees are a potential risk and therefore developed procedures to make it as hard as possible for any one person to recover patient identities. Staff and subcontractors also have to sign contracts prohibiting attempts to re-identify patients.

Such legal and procedural measures are essential to maintaining anonymization, but, like any form of security, they cannot give absolute protection against re-identification. "No, anonymization doesn't really protect privacy", Anderson said about patients' medical records even if personal information such as name or birth date is removed. As an example he cited the date on which Tony Blair, the UK's prime minister at the time, was treated in Hammersmith hospital for atrial fibrillation, which could be readily identified given the publicity over the event. These issues have been recognized by leading advocates, such as the International Pharmaceutical Privacy Consortium (IPPC), which stressed the need to make anonymization deeper than just removing names, post-codes and dates of birth and extend it to biometric details and device identifiers or admission and discharge dates (http://pharmaprivacy.org/assets/activities/IPPC_White_Paper_Anonymisation_Clinical_Trials_Data.pdf).

Just like security, anonymization or pseudonymization requires constant updates to ensure that health data sets are robust against attempts to re-identify the individuals who provided them. Pseudonymization is therefore not defined by technical standards but by the outcome, or the level of difficulty involved in re-identification, according to Peteris Zilgalvis, Head of eHealth and Well Being at the European Commission's Directorate General for Communications Networks. "We can't be locked into just removing names from biological materials which might not lead to a person's identify today but might do so in 10 or 20 years' time or even sooner", he said.

In addition, there is the complex issue of data ownership. This is rather straightforward when individuals collect health data by their own choice, such as monitoring heart rates during exercise; people have total ownership and full rights to delete. "But there is the question of when someone has a medical intervention, does that belong to them, to the surgeon, the nurse, the hospital, or the insurer who paid for it", Zilgarvis commented. "It's a little difficult to say who has ownership in the sense of exclusive use". This may be clearer in cases of serious infectious diseases where there is an obvious public interest in retaining records and therefore prohibiting individuals from erasing data. But there is still a need to determine who has rights of access. "We are in the process of discussing that and how to frame the regulations", Zilgarvis said. "We will probably have a Commission document on this out this calendar year setting out some of the issues like data portability or geographical restrictions". He anticipates that data rights and access would be governed by context so that for example public health authorities would have access to relevant data in the event of an epidemic. "Similarly, auditing authorities would need to see data when checking for fraud or wastage", Zilgarvis added. While the trend goes towards greater harmonization of such rules on medical data ownership, according to Zilgarvis, there are still divergent opinions over access to data for research purposes. "There tend to be more opportunities for using epidemiological data within the Nordic countries and the UK and correspondingly less further south", Zilgarvis said.

These challenges will be further exacerbated with the increasing implementation of Mobile Health (mHealth). One question regards the quality of data produced by a range of diagnostic devices, health-related apps for smartphones and other mobile devices, such as fitness monitors and eWatches, while another concerns the impact that new healthcare players such as Google and Apple may have. "There's many different kinds of data, some sensitive and some non-sensitive but still personal, and some that can be in the public domain, which all needs treating in different ways", Zilgarvis commented. In an attempt to provide some early guidance, the EC published a draft document (https://ec.europa.eu/digital-single-market/en/news/current-initiatives-unlock-potential-mobile-health-europe) to provide a basic framework with the objective of ushering in controlled use of mobile technology for the benefit of health care.

"As the collection of personal health data continues to proliferate both in volume and scope, the biggest requirement is greater transparency and clarity over how the data is being used..."

Security in the sense of protecting against unauthorized access, as opposed to anonymizing to hide identities, is an issue of growing importance as mHealth data becomes more widely distributed. The Commission has in fact ranked privacy and security as the highest priorities in order to gain the trust of citizens from the outset. The key tenet is that the proliferating range of mHealth apps should comply with established EU data protection rules.

Data quality and integrity are also issues that have yet to be addressed properly, according to Peter Doshi, assistant professor of pharmaceutical health services research in the School of Pharmacy, University of Maryland, USA: "I in general have major concerns about data quality when it comes to these areas such as mobile health and not one I think that is easily addressable either at the side of data collection, or later during analysis". Ensuring the quality of data that health apps collect and process will be essential if such apps have access to electronic health records and are incorporated in clinical practice. The European Commission is inviting proposals for assuring data quality in mHealth and has set up a working group to study the issue.

As the collection of personal health data continues to proliferate both in volume and scope, the biggest requirement is greater transparency and clarity over how the data is being used, according to Adam Tanner at the Institute for Quantitative Social Science at Harvard University, USA, and author of several books on the subject. "The system as it is evolving threatens to undermine the trust of patients as they begin to understand the commercial trade in their medical secrets", he said. "Telling patients in clear language what is going on and giving them a choice whether to participate in information sharing will help ensure that scientists gain access to data vital to their research. [...] Many patients would happily share health care data if they know it will be used to advance science. But more regulation is likely necessary to prod commercial firms in that direction".